Оглавление

Предисловие

За последнее десятилетие функция поиска проникла во все сферы – поле поиска по ключевой фразе фактически стало неотъемлемой частью пользовательского интерфейса большинства веб-сайтов и приложений. В то же время поддержка по-настоящему релевантного поиска для большинства организаций оставалась слабым местом, порой не выдерживающим никакой критики.

За это время появились мощные технологии с открытым исходным кодом, обеспечивающие быстрый полнотекстовый поиск (Apache Lucene) в распределенном и легко масштабируемом окружении, практически не требующем писать дополнительный код (Apache Solr и позднее Elasticsearch). Они обеспечили необходимую инфраструктуру для создания «в целом релевантных» механизмов поиска в масштабе реального времени, соответствующих требованиям эпохи больших данных. По мере устранения проблем, характерных для инфраструктур жесткого поиска, и стандартизации решений, многие организации стали отказываться от масштабируемых механизмов быстрого поиска в пользу инструментов, гарантирующих более релевантное соответствие информационным потребностям пользователя. Иными словами, предоставление «в целом релевантных» результатов уже не является достаточным – Google и другие крупные поисковые системы уже приучили пользователей, что поисковые приложения почти читают их мысли. В этой книге рассказывается, как еще активнее двигаться в направлении понимания намерений пользователей.

Даг Тарнбулл (Doug Turnbull) и Джон Берримен (John Berryman) – два опытных эксперта в области технологий релевантного поиска. Мы знакомы уже много лет и часто встречаемся на конференциях, посвященных проблемам поиска. Я с удовольствием вспоминаю время, проведенное с ними за обсуждением идей решения самых сложных проблем, связанных с релевантностью поиска, персонализацией и автоматическим подбором рекомендаций. Никто так не волновался, ожидая знакомства с их опытом, облеченным в форму этой книги – одной из лучших технических книг, которые я когда-либо читал.

Настройка релевантности – сложная проблема. Многие неправильно ее понимают, и часто это остается неочевидным, пока не обнаружится ошибка. Обычно, чтобы заметить несоответствие требуется увидеть много неудачных примеров, и часто бывает сложно понять, как должны выглядеть удачные результаты, не имея их перед глазами. К сожалению, это редко происходит до внедрения системы в эксплуатацию, когда организации начинают замечать разрыв между желаемыми и фактическими результатами поиска.

Кроме того, навыки (знание предметной области и характерных признаков, знакомство с приемами машинного обучения, онтологии и обработки

естественного языка), необходимые для реализации релевантного поиска, сильно отличаются от тех, что требуются для создания и сопровождения масштабируемых инфраструктур (распределенные системы, структуры данных, производительность и конкурентное выполнение, знание аппаратного обеспечения, приемы организации сетевых взаимодействий). Должность специалиста по релевантности вообще отсутствует во многих организациях, из-за чего остается нереализованным потенциал релевантного поиска, который по-настоящему радует пользователей и помогает двигать компанию вперед.

Спектр персонализации между поиском по ключевым фразам, вводимым вручную, и автоматическим подбором рекомендаций также открывает широкое поле деятельности для внедрения релевантности и обеспечения более полного соответствия потребностям пользователей. Авторы проделали большую работу, объясняя некоторые из тонкостей, связанных с поиском, позволяющих в полной мере использовать этот спектр. Вооруженные приемами, описываемыми в этой книге, вы сможете взять на себя роль специалиста по релевантности и найти решение большинства проблем, присущих системам настоящего релевантного и персонализированного поиска.

Трей Грейнджер (Trey Grainger)
Автор книги «Solr in Action»
Первый вице-президент компании
«Engineering at Lucidworks»

Вступление

Джон и я познакомились, когда работали вместе как консультанты в Open-Source Connections (OSC) и решали проблемы поиска для клиентов. Для одних мы искали более производительные решения (увеличивали скорость работы!). Другим помогали в создании поисковых приложений. Все наши проекты имели простую меру успеха. Удалось увеличить производительность? Приложение было закончено?

Однако релевантность поиска – совсем другое дело. Пользователи, взращенные в эпоху Google, не хотят иметь «просто хороший» поиск. Им нужен «чертовски умный» поиск. Они желают иметь возможность определять критерии поиска, и их не устраивает, когда поисковая система подбирает результаты «наобум».

Как мотыльки, привлекаемые пламенем, мы оба чувствовали тягу к этой трудной проблеме. И так же как мотыльки, мы часто обжигали свои крылья. Усваивая эти болезненные уроки, мы неуклонно росли и набирались опыта, и теперь с успехом решаем задачи, которые раньше казались нам слишком сложными.

За это время окрепли наши голоса в блоге OSC. Мы обнаружили, что о проблемах релевантного поиска пишется непозволительно мало. Мы выдвинули и реализовали идею релевантности, управляемой тестированием. Мы подробно описали наши трудности, проблемы и победы. Вместе мы экспериментировали с приемами машинного обучения, такими как латентно-семантический анализ. Мы углубились в изучение внутренних механизмов Lucene и исследовали приемы создания компонентов поиска для решения проблем. Мы начали изучение темы поиска информации. Узнавая больше приемов решения сложных проблем, мы тут же писали о них.

Но блоги имеют свои ограничения. Мы с Джоном всегда хотели выразить наши идеи более системно, в форме книги. К счастью, произошла целая цепочка благоприятных событий, благодаря которым у нас появилась такая возможность. Я, вместе с Эндрю Монталенти (Andrew Montalenti), выступил с докладом о средствах параллельного выполнения в Python на местной встрече. Эндрю упомянул этот доклад на конференции PyCon и в результате ему позвонили из издательства Manning, чтобы обсудить возможность создания книги о параллельном программировании на Python. Эндрю ответил, что не собирается писать книгу, но его содокладчик Даг, возможно, согласится.

Так получилось, что я тоже не особенно хотел писать книгу о параллельном программировании на Python, но подкинул идею написать другую книгу. Я подошел с этой идеей к Джону и, побеседовав с ним на эту тему два-три раза, мы подготовили для издательства предложение, от которого они не смогли отказаться!

Этот важный звонок из Manning случился почти два года тому назад. С тех пор много воды утекло. За время, пока мы писали книгу, у нас обоих в семьях произошло прибавление. Я начал практиковать оказание консультационных услуг о релевантности. Джон поменял работу, став экспертом по поиску в Eventbrite. И все же мы не могли сопротивляться желанию написать об этой потрясающей теме.

Эта книга не похожа на другие технические книги. Она не является перечислением особенностей одной технологии. Скорее это карта с нашим маршрутом через многие годы боли, и решений проблем, на которые не было готовых ответов. Иными словами, мы прошли насквозь пустыню релевантного поиска, встретили множество оазисов и научились уклоняться от встреч с песчаными людьми и штурмовиками.

Мы передаем вам эту карту, чтобы вы не блуждали в пустыне, как это случилось с нами. А теперь извините, но мы собираемся отправиться на ближайший пляж и отдохнуть…

Даг Тарнбулл

Благодарности

За несколько недель до начала работы над этой книгой, в наших семьях произошло прибавление. Мы выражаем самую глубокую благодарность нашим супругам, Харе Тарнбулл (Khara Turnbull) и Кумико Берримен (Kumiko Berryman). Они много недель не имели выходных, пока мы работали над книгой – за это время Хара закончила свою книгу, а Кумико занималась продажей дома и организацией переезда в другую страну. Настало время для долгого отпуска!

Эта книга не состоялась бы без основателя OpenSource Connections – Эрика Пью (Eric Pugh). Как наш «босс», он подталкивал нас писать, говорить и решать большие проблемы. Как настоящий лидер, Эрик делает свою страсть вашей страстью. Если бы Эрик не отпускал нас в свободное плавание (а иногда даже не подталкивал), мы не смогли бы поверить в свои силы и возможности, как писатели и специалисты, способные решать сложные проблемы. Эрик учил нас, что всякий, в том числе и мы, может стать лидером.

Спасибо проекту TMDb за предоставленные данные и поддержку. Мы потратили массу времени, пытаясь подобрать хорошие наборы данных. Проект TMDb (http://themoviedb.org) не только предоставил богатый набор поисковых данных, но также поддерживал нас и наших первых читателей, пока мы выявляли и исправляли ошибки и неточности, в основном в нашем собственном коде. Тревис Белл (Travis Bell) заслуживает особой нашей благодарности за оперативную реакцию на наши проблемы и электронные письма.

Создание книг – это командный вид спорта и мы хотели бы поблагодарить всех сотрудников издательства Manning, участвовавших в работе над этой книгой: Марину Майклз (Marina Michaels), нашего редактора-консультанта по аудитории; Аарона Колкорда (Aaron Colcord), технического редактора; Валентина Креттаза (Valentin Crettaz), технического корректора; Френка Полманна (Frank Pohlmann) и Майка Стивенса (Mike Stephens), рецензентов издательства; и Кэндис Джиллхули (Candace Gillhoolley) за маркетинг.

Мы также хотим сказать спасибо многим рецензентам, читавшим первые рукописи книги и давшим ценные советы, включая Джона Гатри (John Guthrie), Мартина Бира (Martin Beer), Артура Зубарева (Arthur Zubarev), Эльмана Кринкера (Elman Krinker), Амита Ламба (Amit Lamba), Марка-Оливера Шиила (Marc-Oliver Scheele), Яна Стерка (Ian Stirk), Джозефа Вана (Joseph Wang), Стюарта Вудворда (Stuart Woodward), Урсина Стаусса (Ursin Stauss), Расса Кама (Russ Cam), Майкла Финка (Michael Fink), регора Зуровски (Gregor Zurowski), Димитриоса Кузиса-Лукаса (Dimitrios Kouzis-Loukas), Джереми Гейлора (Jeremy Gailor) и Кейт Вебстер (Keith Webster).

Особое спасибо Эндрю Монталенти (Andrew Montalenti), связавшему нас с издательством Manning. Спасибо Шаю Банону (Shay Banon), создателю Elasticsearch, за его поддержку, а еще просто за то, что он отличный парень. Спасибо нашим коллегам: Трею Грейнджеру (Trey Grainger), Мэтту Оверстриту (Matt Overstreet), Рене Морзе (Rena Morse), Дэвиду Смайли (David Smiley), Гранту Ингерсоллу (Grant Ingersoll), Йонику Сили (Yonik Seeley), Рене Криглеру (Rene Kriegler), Питеру Диксону-Моисею (Peter Dixon-Moses), Чарли Халлу (Charlie Hull) и Дрю Фаррису (Drew Farris) за постоянный обмен мнениями на тему поиска и релевантности на протяжении многих лет. И особое спасибо Трею за предисловие к этой книге.

Спасибо всем членам наших семей за их поддержку. Особенно нашим детям: Мегуме Берримен (Megume Berryman), Яну Тарнбуллу (Ian Turnbull) и Мюррею Тарнбуллу (Murray Turnbull). Спасибо нашим «семьям на работе» в OpenSource Connections и Eventbrite, что позволили нам вложить всю нашу энергию в эту книгу.

Об этой книге

Книга «*Релевантный поиск*» рассказывает, как отвечать на запросы пользователей содержимым, удовлетворяющим их потребности. Она научит вас строго контролировать ранжирование результатов поиска на основе четких критериев, а не мистических прихотей поискового движка. Мы изложим наш подход к глубокой настройке релевантности в Solr или Elasticsearch, а также познакомим с методами, которые помогут вам понять, что является релевантным для вашего приложения.

Кому адресована книга

Книга «Релевантный поиск» предназначена для разработчиков, использующих Solr или Elasticsearch, столкнувшихся с проблемой, когда поисковый механизм возвращает пользователям не всегда релевантные результаты поиска. Читатели, имеющие хотя бы поверхностное знакомство с их поисковым механизмом, смогут с помощью этой книги подняться на следующий уровень. Несмотря на то, что эта книга относится к разряду технических, значительная ее часть посвящена организационным стратегиям – для руководителей проектов, контент-стратегам, специалистам по маркетингу или предметной области, занимающимся проблемами поиска.

Краткое содержание

Книга «Релевантный поиск» постепенно раскрывает технические детали стратегии преодоления проблем, с которыми вам придется столкнуться при определении и решении задачи релевантного поиска. Книга завершается следующими шагами: как организовать персонализированный поиск, семантический поиск и автоматический подбор рекомендаций.

Глава 1 начинается с обсуждения проблемы релевантности. Она охватывает такие области, как веб-поиск, электронная коммерция и экспертный поиск. Глава обсуждает степень поддержки академическими кругами наших стремлений к релевантному поиску. В конце мы изложим техническую стратегию поддержки релевантности.

Глава 2 представляет краткий обзор основных структур данных и алгоритмов в Lucene, имеющих отношение к релевантности. Вы увидите, как Lucene обеспечивает невероятную основу для организации релевантного поиска.

Глава 3 научит приемам отладки релевантного поиска. Когда структуры данных и алгоритмы, обсуждаемые в главе 2, не дают желаемых результатов, вам придется копнуть глубже, чтобы понять, что мешает поиску.

Глава 4 покажет, как разложить содержимое и поиск на описательные признаки, используя процесс анализа в поисковом механизме. Это фун-

даментальное умение поможет вам использовать анализ для поиска чего угодно.

Глава 5 начинает обсуждение стратегий запросов по нескольким полям. В этой главе мы расскажем, как конструировать запросы, оценивающие конкретные факторы ранжирования, важные для ваших пользователей.

Глава 6 продолжает обсуждение стратегий запросов. Здесь мы сосредоточимся на стратегиях поиска по ключевым терминам, которые поддерживают упрощенное понимание релевантности пользователями.

Глава 7 демонстрирует такие приемы регулирования, как форсирование и фильтрация. Вам часто придется управлять поиском, отдавая приоритет более свежей информации, прибыльным продуктам или близлежащим местоположениям.

Глава 8 показывает альтернативные пути, ведущие пользователей к релевантной информации. Иногда, когда ранжирование по релевантности терпит неудачу, направить пользователя в правильном направлении могут помочь некоторые особенности пользовательского интерфейса, такие как категории, автодополнение и выделение.

Глава 9 описывает создание приложения с полноценной поддержкой релевантного поиска, что поможет вам глубже вникнуть в суть. Получив достаточно полное представление о релевантном поиске в предыдущих главах, в этой главе вы увидите весь процесс создания продукта, от начала до конца.

Глава 10 поднимается на ступень выше в исследовании стратегий создания продукта и рассматривает культурные и организационные факторы. Какая информация считается релевантной для организации? Здесь вы увидите, как организация может реализовать быструю и точную обратную связь для направления усилий специалиста по релевантности.

Глава 11 выведет вас за рамки поискового механизма. Здесь вы узнаете, как можно улучшить релевантность поиска за счет применения приемов машинного обучения, персонализации и семантического поиска.

Приложение A подробно, шаг за шагом, описывает процесс загрузки данных для примеров из книги в Elasticsearch посредством The Movie Database (TMDB) API.

Приложение B предназначено для знатоков Solr и описывает соответствие функций поддержки релевантности в Elasticsearch и Solr.

Об исходном коде

Эта книга содержит много примеров исходного кода, и в виде листингов, и в виде фрагментов в обычном тексте. В обоих случаях исходный код оформляется моноширинным шрифтом, чтобы его можно было отличить от обычного текста. Иногда, чтобы подчеркнуть отличия от предыдущего шага и

выделить вновь добавленные особенности, код будет оформляться **жирным моноширинным шрифтом.**

Во многих случаях оригинальный исходный код был переформатирован; мы добавили переносы строк и изменили ширину отступов, чтобы уместить строки кода по ширине книжной страницы. Кроме того, мы убрали комментарии из кода, если он описывается в тексте книги. Многие листинги сопровождаются дополнительными аннотациями, подчеркивающими наиболее важные идеи.

Примеры были проверены с Elasticsearch 2.0 и Python 2.7.

Исходный код примеров для глав 3–9 можно найти на веб-сайте издательства Manning (www.manning.com/books/relevant-search) и репозитории книги на сайте GitHub (http://github.com/o19s/relevant-search-book). Примеры создавались в среде iPython Notebook/Jupyter, чтобы упростить эксперименты с ними. В файле README вы найдете подробный перечень дополнительного программного обеспечения, которое требуется установить для опробования примеров.

Автор в сети

Одновременно с покупкой книги «Релевантный поиск» вы получаете бесплатный доступ к частному веб-форуму, организованному издательством Manning Publications, где можно оставлять комментарии о книге, задавать технические вопросы, а также получать помощь от автора и других пользователей. Чтобы получить доступ к форуму и зарегистрироваться на нем, откройте в веб-браузере страницу www.manning.com/relevant-search. Здесь описывается, как попасть на форум после регистрации, какие виды помощи доступны и правила поведения на форуме.

Издательство Manning обязуется предоставить своим читателям место встречи, где может состояться содержательный диалог между отдельными читателями и между читателями и автором. Но со стороны автора отсутствуют какие-либо обязательства уделять форуму какое-то определенное внимание – его присутствие на форуме остается добровольным (и неоплачиваемым). Мы предлагаем задавать автору стимулирующие вопросы, чтобы его интерес не угасал!

Форум и архив с предыдущими обсуждениями остается доступным на сайте издательства, пока книга продолжает издаваться.

Другие онлайн-ресурсы

Желающим узнать больше мы рекомендуем обращаться к следующим замечательным ресурсам:

- блог OpenSource Connection (http://opensourceconnections.com/blog);
- личный блог Джона Берримена (John Berryman, http://thoughtbox.solutions);

- блог проекта Elastic (www.elastic.co/blog);
- блог проекта Lucidwork (https://lucidworks.com/blog);
- блог «Salmon Run» Суджита Пала, посвященный Solr (http://sujitpal.blogspot.com/);
- новостная рассылка Solr Start (www.solr-start.com).

Желающим изучить тему поиска и извлечения информации в целом, мы рекомендуем следующий канонический труд:

- «Introduction to Information Retrieval» Кристофера Маннинга (Christopher Manning) с соавторами. (Cambridge University Press, 2008), http://nlp.stanford.edu/IR-book/.[1]

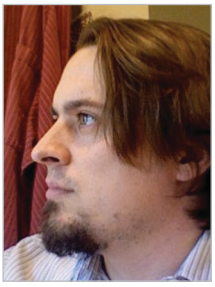
С вопросами, касающимися конкретно технологий Solr/Elasticsearch, мы рекомендуем обращаться на специализированные дискуссионные форумы:

- по теме Elasticsearch: http://discuss.elastic.co;
- по теме Solr: http://lucene.apache.org/solr/resources.html.

[1] *Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце*. Введение в информационный поиск. М:. Вильямс. 2014. ISBN 978-5-8459-1623-5, 978-0-5218-6571-5. – *Прим. перев.*

Об авторах

Даг Тарнбулл (Doug Turnbull) ведет консультационную практику в OpenSource Connections, где часто выступает и ведет свой блог. Даг конструирует системы релевантного и семантического поиска для своих клиентов в самых разных областях, используя разнообразные технологии поиска и обработки естественного языка.



Джон Берримен (John Berryman) начинал карьеру как инженер в аэрокосмической промышленности, но спустя несколько лет обнаружил, что ему больше нравится программировать или решать сложные математические задачи. В конце концов Джон забросил самолеты и спутники, и занялся разработкой программного обеспечения, программных архитектур и поисковых технологий. В настоящее время Джон работает в Eventbrite, помогая создавать системы обнаружения событий, поиска и автоматического подбора рекомендаций с применением Elasticsearch.

Об иллюстрации на обложке

На обложке книги изображен рисунок, подписанный как «Homme de l'Isle de Pathmos», или человек с острова Патмос (Греция). Иллюстрация взята из каталога костюмов разных стран, составленного Жаком Грассетом де Сен-Совер (Jacques Grasset de Saint-Sauveur, 1757–1810), озаглавленного как «Costumes de Différents Pays» и опубликованного во Франции в 1797. Каждая иллюстрация в этом каталоге нарисована и раскрашена вручную. Многообразие рисунков в каталоге Грассета де Сен-Совер отчетливо демонстрирует уникальные и индивидуальные особенности городов и районов мира, существовавших 200 лет назад. Изолированные друг от друга, люди говорили на разных языках и диалектах. Встретив человека на улице, по его одежде легко можно было определить, где он живет, кем работает и какое положение в обществе занимает.

С тех пор мода изменилась и региональные различия, такие существенные в те времена, исчезли. Сейчас зачастую сложно отличить жителей разных континентов, не говоря уже о жителях из разных городов, областей или стран. Но, если посмотреть на это с оптимистической точки зрения, мы обменяли культурное и визуальное разнообразие на более разнообразную личную жизнь. Или более разнообразную и интересную интеллектуальную жизнь и техническую вооруженность.

Мы в издательстве Manning славим изобретательность, предприимчивость и радость компьютерного бизнеса обложками книг, изображающими богатство региональных различий двухвековой давности, возвращая к жизни иллюстрации из каталога Грассета де Сен-Совер.

Глава 1

Задача релевантного поиска

Эта глава охватывает следующие темы:

- вездесущий поиск (поиск вокруг нас!);
- сложности реализации релевантного поиска;
- примеры сложностей поиска в разных предметных областях;
- неспособность готовых решений справиться с проблемой;
- подход к организации релевантного поиска, освещаемый в этой книге.

Создание механизма поиска, обладающего желаемым поведением, может свести с ума. И только начинающие осваивать Solr или Elasticsearch, и имеющие многолетний опыт использования этих инструментов, рано или поздно сталкиваются с низким качеством результатов поиска. Стандартные настройки часто не соответствуют нашим потребностям, и нам приходится бороться даже за мало-мальски релевантные результаты поиска.

Когда дело доходит до оценки релевантности результатов, механизм поиска может показаться мистическим черным ящиком. Часто возникает желание игнорировать проблемы релевантности и обратить свое внимание на другие, менее мистические, аспекты приложения, такие как производительность или оформление пользовательского интерфейса. К сожалению, избежать оценки релевантности не удастся. Современным пользователям все чаще приходится работать с большими объемами информации, будь то каталоги продуктов, книги, сообщения в файлах журналов, электронные письма, каталоги гостиниц или медицинские статьи. Поле поиска – первое место, куда обращаются пользователи, чтобы приступить к изучению и найти ответы. Не имея возможности получать интуитивные ответы на вопросы, пользователи чувствуют себя потерявшимися. Поэтому, несмотря на сумасшедший и почти мистический характер работы механизмов поиска, вы должны найти решение.

Книга «Релевантный поиск» срывает покров тайны с релевантности. Что такое релевантность на самом деле? Это основная ценность, которую может предложить механизм поиска. *Релевантность* – это искусство определения в процессе поиска ценности информации для пользователя и бизнеса. Дьявол, как известно, кроется в деталях. Какая информация оценивается? (Твиты? Продукты? Детские игрушки?) К какой категории относятся пользователи? (Врачи? Технически подкованные покупатели?) Что является предметом поиска? (Название на японском? Товары известных производителей? Юридический термин?) Что ожидают получить пользователи в ответ? (Список товаров в магазине? Каталог книг в библиотеке?) И что надеется получить от этого взаимодействия ваш работодатель? (Прибыль? Увеличение количества просмотров? Восхищенные отзывы?) Поиск все чаще становится неотъемлемой частью наших приложений, уверенно отвоевывая себе место под солнцем. Ответы на эти вопросы (в отношении релевантности) обусловливают разницу между благоприятными впечатлениями пользователей и их разочарованием.

1.1. Ваша цель: стать специалистом по релевантности

Как ее достичь? Книга «Релевантный поиск» обучит вас всему, что должен знать специалист по релевантности. Обладая этими знаниями, вы сможете трансформировать механизм поиска в интеллектуальную систему, которая понимает потребности пользователей и бизнеса. Для этого вы подскажете механизму поиска, какие характеристики в вашей информации считаются наиболее важными: местоположение ресторана, слова в тексте книги или цвет рубашки. Определив перечень важных признаков, вы сможете оценивать важность информации для пользователей при поиске: как далеко находится ресторан? Эта книга описывает нужную тему? Будет ли эта рубашка соответствовать только что купленным брюкам? Все эти факторы, оцениваемые в процессе поиска и сообщающие о важности или неважности информации для пользователя, называются *сигналами*. Вы увидите, что основной проблемой является выбор признаков и реализация сигналов, отражающих потребности ваших пользователей и бизнеса.

Но техническое мастерство – лишь часть работы (как показано на рис. 1.1). Понимание, что именно требуется реализовать, порой важнее, чем знание, как это сделать. Как ни странно, специалист по релевантности редко знает, что означает «релевантность» для данного приложения. Но этим знанием обладают другие – обычно нетехнические специалисты – понимающие назначение информации и цели пользователей и бизнеса. В этой книге вы научитесь добиваться релевантности, используя знания и опыт других, а также данные о поведении пользователей в процессе поиска.

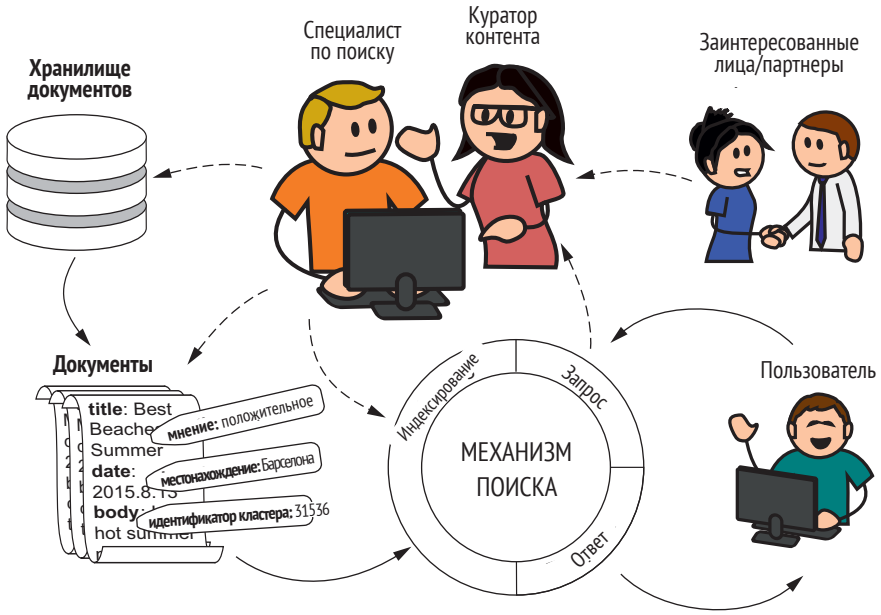


Рис. 1.1. Специалист по релевантности использует механизм поиска и вспомогательные технологии, чтобы выразить бизнес-логику оценки результатов. Он тесно сотрудничает с другими специалистами, способными оценить релевантность, и широко использует информацию о пользователях.

Мы еще будем уточнять упомянутые идеи далее в этой главе (и книге). Но, чтобы заложить надежный фундамент, в оставшейся части этой главы мы определим основные наши задачи и проблемы. Почему релевантный поиск сопряжен с большими сложностями? Что можно предпринять, чтобы справиться с этой сложностью? Затем мы переключимся на обсуждение подхода к решению проблемы релевантности, являющегося основной темой этой книги.

1.2. Сложности релевантного поиска

Сложность реализации релевантного поиска отчасти объясняется тем, что мы принимаем непосредственное участие в акте поиска. Приложения, осуществляющие поиск, получают от пользователей запросы (текст, который они вводят в поле поиска) и пытаются оценить степень соответствия найденной информации.

Этот акт происходит так часто, что остается почти незамеченным. Возьмите в качестве примера самого себя. Возможно вы, проснувшись сегодня утром, включили кофеварку и начали возиться со своим смартфоном. Вы просмотрели новости, заглянули в Facebook и проверили электронную почту. Еще до того, как кофе был готов, вы наверняка успели поработать с десятком поисковых приложений, даже не задумываясь об этом. Вы от-

правили сообщение другу, которого сначала отыскали в списке контактов? Нашли важное электронное письмо? Пообщались с Сири? Удовлетворили свое любопытство, выполнив поиск в Google? Покрутились вокруг телевизора мечты с 50-дюймовым экраном на Amazon?

За короткое время вы испытали продукты, на которые были затрачены тысячи человеко-часов опытных инженеров. Воспользовались результатом продолжительных академических исследований, уходящих корнями в область информационного поиска, зародившуюся в прошлом столетии. Стоя на плечах гигантов, вы просеяли гигантский объем информации – почти все, что было накоплено человечеством по этой теме – и нашли лучший обзор самого популярного телевизора за минуты.

Или, может быть, ваш опыт поиска не был таким радужным. Вполне вероятно, что вы получили разочаровывающие результаты. Может быть вы не смогли найти нужный контакт в своем телефоне, потому что просто допустили опечатку. Возможно механизм поиска не проникся вашей идеей о телевизоре мечты. В отчаянии вы сдались, удалив приложение с мыслью: «Почему так сложно найти что-то осмысленное?»

В действительности для реализации «простого» поиска, который кажется пользователям «осмысленным», часто требуются значительные усилия инженеров. Пользователи возлагают большие надежды на поисковые приложения. Нашим приложениям предлагается мгновенно понять, какую информацию желают получить пользователи, опираясь на несколько запросов, введенных в спешке. Хуже того, пользователям обычно не хватает времени просмотреть пару десятков результатов. Они делают несколько мимолетных попыток и быстро разочаровываются, если, как им кажется, поиск не принес желаемых результатов. Вам дается минимальное время, чтобы успеть вернуть релевантные результаты, и это время постоянно сокращается.

Вы могли бы подумать: «Конечно, задача выглядит сложной, но почему ее так нелегко решить?» Проблема поиска существует уже довольно давно; разве этого недостаточно, чтобы научить поисковые механизмы, такие как Solr или Elasticsearch, всегда возвращать правильные результаты? Или, почему бы просто не отправлять пользователей в Google? Почему стандартные коммерческие решения, такие как Amazon A9, не способны решить наши проблемы с поиском?

1.2.1. Какой результат можно назвать «релевантным»?

Легко обмануться, рассматривая поиск, как отдельную задачу, изолированную от всего остального. В действительности поисковые приложения сильно отличаются друг от друга. Да, верно, типичное приложение поиска дает пользователю возможность вводить текст, фильтровать документы и взаимодействовать со списком ранжированных результатов. Но пусть вас не обманывают внешние атрибуты. Все приложения имеют совершенно

разные представления о релевантности. Давайте рассмотрим несколько типичных классов поисковых приложений, чтобы вы могли ощутить, насколько уникально определение релевантности для вашего приложения.

Сначала рассмотрим веб-поиск. С ростом Всемирной паутины все больше стало появляться сомнительных сайтов, легко обманывающих ранние поисковые системы. Создатели таких сайтов вставляли в свои страницы фразы, чтобы обмануть поисковую систему. В лучшем случае ранние поисковые системы возвращали любые старые совпадения с запросом пользователя. В худшем, они приводили пользователей на страницы со спамом или даже с вирусами.

В Google быстро осознали, что релевантность веб-поиска зависит не только от текста, но и от доверия. Пользователям нужно было помочь отсеять весь этот мусор. Поэтому в Google разработали алгоритм PageRank[1] для оценки степени доверия к содержимому. PageRank вычисляет показатель доверия, определяя количество внешних веб-ссылок на сайт. Благодаря PageRank, Google возвращает не только содержимое, соответствующее запросу пользователя, но также содержимое, которое выглядит надежным и вызывает доверие у остальной части Всемирной паутины. Этот акцент на возврате надежной информации продолжает играть важную роль и в наше время, когда Google стремится поставить заслон перед вредоносными веб-сайтами, которые в свою очередь не ослабляют попыток обмануть поисковую систему.

Теперь сравним веб-поиск с электронной коммерцией. Такие сайты, как Amazon, полностью контролирующие хранящуюся на них информацию, не испытывают проблем с доверием. Понятие релевантности для пользователей этих сайтов имеет тот же смысл, что для любых покупателей: поиск должен возвращать доступные по цене и высококачественные товары. Но сайты электронной коммерции интересуют не только покупатели. У них есть свои интересы. Они должны возвращать результаты, генерирующие прибыль, помогающие распродать залежавшийся товар и удовлетворяющие требованиям производителей.

Функция поиска на сайте электронной коммерции одновременно играет роль продавца. Специалист по релевантности должен запрограммировать в функцию поиска те же приоритеты, которые имеют значение для любого торгового предприятия. Специалист по релевантности должен создать такую процедуру поиска, чтобы она понимала желания покупателей, а покупатели покидали онлайн-магазин удовлетворенными и с покупками. Релевантными для сайта электронной коммерции считаются результаты поиска, которые не только помогают покупателям сделать покупки, но и приносят прибыль.

[1] За подробностями обращайтесь к статье «The Anatomy of a Large-Scale Hypertextual Web Search Engine» Сергея Брина (Sergey Brin) и Лоуренса Пейджа (Lawrence Page): http://infolab.stanford.edu/~backrub/google.html.

Другие виды поиска, особенно в медицине, юриспруденции и науке, должны исследовать содержимое документов для определения их релевантности. Такой экспертный поиск зависит от понимания профессионального жаргона, вводимого специалистами, например, юристами или врачами. Эти решения должны понимать тонкости предметной области – например, что «сердечный приступ» – это то же самое, что «инфаркт миокарда». Или, что острый «инфаркт миокарда» – это особый тип «сердечного приступа».

Подобно тому, как процедура поиска онлайн-магазина отражает взаимодействие продавца с покупателем, экспертный поиск напоминает диалог ученого с опытным библиотекарем. Библиотекарь понимает профессиональный сленг ученого. Отвечая на вопрос, он отправляет его к данным и схожим исследованиям, которые ученому трудно было бы найти самостоятельно.

Определение релевантности для таких поисковых приложений зависит от решений, первоначально предназначавшихся для организации информации в библиотеках. Например, в медицине, рубрикатор MeSH (Medical Subject Headings – медицинские предметные рубрики), изображенный на рис. 1.2, располагает темы в порядке от менее конкретных к более конкретным и упрощает извлечение информации по синонимам. Для экспертного поиска *релевантность* подразумевает тщательное связывание тем запросов и ответов. Релевантным считается результат, который может вызвать у ученого восклицание «Ага!» и внезапное понимание, чего он не смог бы так легко найти самостоятельно.

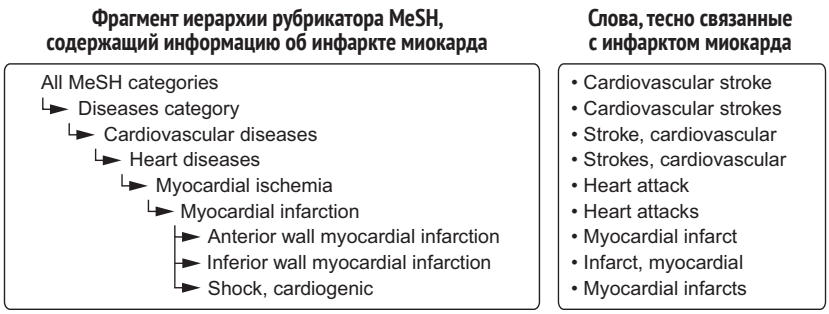


Рис. 1.2. Раздел «Myocardial Infarction» (инфаркт миокарда) в рубрикаторе MeSH (слева) и некоторые темы в MeSH, тесно связанные с этим заболеванием

1.2.2. Поиск: не существует панацеи от всех бед!

Классы проблем, которые мы только что обсудили, – это лишь вершина айсберга удивительного разнообразия поиска. Можно ли считать поиск недвижимости разновидностью поиска в онлайн-магазине? Некоторое сходство конечно есть (удовлетворенные пользователи делают покупки), но для

покупающего дом большое значение приобретает еще целый ряд факторов (наличие школ поблизости, экология, количество спален). А что можно сказать о поиске ресторанов? Или продуктов? О заказе блюд в ресторане? О поиске добровольных помощников? Или о поиске кого-то, кто будет чистить снег после снегопада? А как насчет поиска в локальной сети? А что вы скажете о своем приложении? Как вы определяете релевантность?

Учитывая огромное разнообразие требований к релевантности, удивляет, как много существует производителей, обещающих поставить настоящую панацею от всех бед. Ваше определение релевантности, скорее всего, намного более уникальное, чем вы думаете. У ваших пользователей имеются ожидания, о которых они даже не подозревают. Ваше информационное наполнение и бизнес несут проблемы, с которыми вы еще не сталкивались.

В самом деле, будьте благодарны, что Solr или Elasticsearch со стандартными настройками не удовлетворяют вашим потребностям. Вы выбирали язык программирования не потому, что ваш продукт является лишь модулем его стандартной библиотеки. Если бы это было так, в вашем продукте не было бы ничего уникального! Думайте о Solr или Elasticsearch, как о программной инфраструктуре для организации поиска. Механизм поиска с открытым исходным кодом позволяет *вам* запрограммировать ваше понимание релевантности. Мы научим вас именно этому: искусству и науке создания релевантных решений с использованием открытых технологий, удовлетворяющих целям пользователей и бизнеса.

1.3. Анализ релевантности

Итак, как видите, ваше приложение имеет собственное определение релевантности. Но почему нет универсальной, четко определенной практики возврата релевантных результатов поиска? Поискав в Интернете, можно найти большое количество штучных решений, прекрасно справляющихся с конкретными проблемами. Создается ощущение, что релевантный поиск не имеет единого, целостного и универсального решения, а вместо этого существует набор разрозненных трюков.

В действительности это не совсем так: в академическом поле есть дисциплина, лежащая в основе релевантности, которая называется информационным поиском. Она предлагает обобщенные практики повышения релевантности во многих областях. Но, как мы уже видели, релевантность сильно зависит от конкретного приложения. Учитывая это и знакомясь с принципами информационного поиска, думайте, как их можно использовать для решения вашей узкоспециальной задачи релевантного поиска.[2]

[2] Для знакомства с темой информационного поиска мы рекомендуем классический труд «Introduction to Information Retrieval» Кристофера Маннинга (Christopher Manning) с соавторами. (Cambridge University Press, 2008), http://nlp.stanford.edu/IR-book/. (*Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце*. Введение в информационный поиск. – М:. Вильямс. 2014. ISBN 978-5-8459-1623-5, 978-0-5218-6571-5. – *Прим. перев.*)

1.3.1. Информационный поиск

К счастью, проблемы поиска изучаются экспертами уже несколько десятилетий. Академическая дисциплина информационного поиска сосредоточена на проблемах точного извлечения информации, удовлетворяющей потребности пользователя. Какая *информация нужна*? Думайте об этом, как о *спецификации* идеального контента, удовлетворяющего информационным потребностям пользователя. Эта спецификация не ограничивается искомой строкой. Например, представьте такую проблему программирования: библиотечная Java-функция sort возбуждает исключение NullPointerException и вам нужно выяснить, почему это происходит. Информационную потребность в этом случае можно определить так:

> Определить, почему вызов метода `sort` в моем приложении вызывает исключение `NullPointerException`. (также было бы неплохо увидеть пример кода, решающего эту проблему, чтобы я смог пойти на обед!)

Чтобы удовлетворить эту информационную потребность, вы, скорее всего, введете строку запроса «sort метод NullPointerException» или «<фрагмент кода> NullPointerException». Если вам повезет, вы найдете решение проблемы, схожей с вашей. Эта информация поможет вам решить вашу проблему и вы сможете продолжить движение вперед.

В информационном поиске *релевантность* определяется как практика возврата результатов поиска, максимально удовлетворяющих информационные потребности пользователя. Кроме того, классический информационный поиск предусматривает ранжирование текста. Многие приемы в информационном поиске пытаются определить, насколько точно та или иная статья соответствует запросу пользователя. С некоторыми из этих бесценных приемов вы познакомитесь в данной книге, так как многие из них реализованы в механизмах поиска с открытым исходным кодом.

Для выявления лучших методов, исследователи в области информационного поиска испытывают разные стратегии, используя тестовые коллекции статей. В эти коллекции входят обзоры с Amazon, новости с сайта информационного агентства Reuters, сообщения в Usenet и другие подобные текстовые данные с размерами, соответствующими статьям. Для помощи в испытаниях решений эти коллекции аннотированы в экспериментальном окружении примечаниями, отмечающими, какие результаты наиболее релевантны для данного запроса. Например, при поиске по строке «Mitt Romney» (Митт Ромни), наиболее релевантными будут считаться новости о его президентских выборах в 2008 и 2012. Статьи о предыдущей работе Ромни, возможно, будут считаться менее релевантными. Статьи, обсуждающие его отца, Джорджа Ромни (George Romney), скорее всего, будут оценены, как еще менее релевантные. Такие аннотированные списки

результатов поиска, релевантных для определенного набора запросов, известны как *оценочные списки* (см. рис. 1.3).

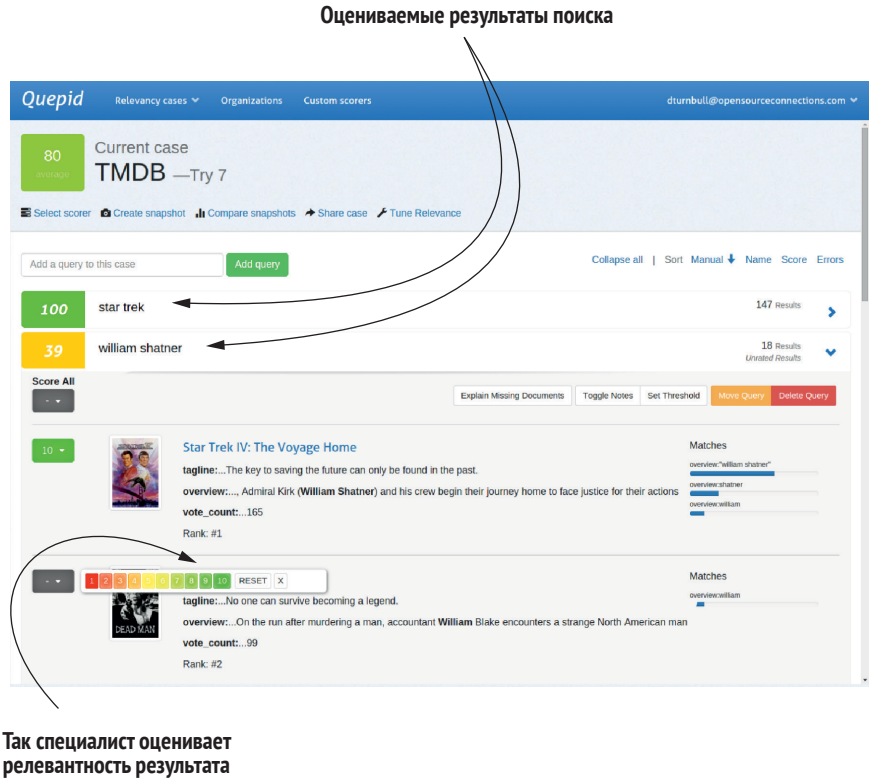


Рис. 1.3. Пример создания оценочного списка для запроса «Rambo» в приложении Quepid управления оценочными списками

Используя оценочные списки, исследователи определяют, какие изменения в процедуре вычисления релевантности, повышают общее соответствие результатов для каждой тестовой коллекции. В классическом информационном поиске успешным считается решение, улучшающее общую релевантность на 1% для десятка коллекций. Вместо решений для каждой конкретной задачи, основное внимание в дисциплине информационного поиска уделяется поиску решений для широкого круга проблем.

1.3.2. Можно ли использовать достижения информационного поиска для решения задачи релевантности?

Вы уже успели убедиться, что нет универсального решения. Но, похоже, что информационный поиск систематически создает релевантные решения. Поэтому спросите себя: применимы ли эти идеи к вашему приложе-

нию? Можно ли использовать в вашем приложении решения, обеспечивающие общее увеличение соответствия для поиска в текстах статей? Может быть лучше решать более конкретные проблемы, возникающие в вашем приложении?

Проще говоря, классический информационный поиск поднимает несколько вопросов, связанных с прикладными проблемами релевантности. Перечислим их, чтобы посмотреть, когда достижения в информационном поиске могут помочь, а когда оказываются бесполезными.

- *Является ли удовлетворение информационных потребностей единственной целью?* Во многих приложениях удовлетворение информационных потребностей пользователей не является единственной целью. Поиск также служит целям бизнеса. Вы видели это выше, на примере с электронной коммерции. Хотя часто говорят «клиент всегда прав», также верно, что компании не могут существовать без продажи рекламы, создания прибыли, удовлетворения производителей и движения товаров. В любом поиске имеются дополнительные факторы, которые ставят потребности бизнеса выше информационных потребностей пользователей. Подобно тому, как продавцы подержанных автомобилей стремятся сбыть товар по завышенной цене, специалисты по релевантности должны учитывать эти факторы, чтобы не привести своего работодателя к разорению.
- *Что еще кроме текста отражает информационные потребности?* В центре внимания классического информационного поиска находится обобщенная (на все случаи жизни) мера релевантности текста. Это обстоятельство может быть неприемлемым – вообще – для вашего приложения, требующего больше внимания уделить конкретным проблемам. Один такой пример мы рассмотрели выше: как в Google усовершенствовали веб-поиск, внедрив числовую меру доверия к веб-сайтам (алгоритм PageRank). Google использует алгоритм PageRank в дополнение к метрикам, основанным исключительно на оценке текстов, которые легко можно подделывать. Но даже текстовый поиск не всегда точно вписывается в задачу информационного поиска, ориентированного на текстовые данные с размерами, соответствующими статьям. Хорошие результаты, полученные для коротких текстовых фрагментов, таких как твиты или заголовки, требуют особого осмысления. Вы, не будучи исследователями в информационном поиске, должны решить, какие факторы имеют значение для вашего приложения, и реализовать их. Решение, дающее плохие результаты на наборе новостей с сайта информационного агентства Reuters, может оказаться именно тем, что вам нужно.
- *Что подразумевается под информационными потребностями пользователя?* Часто обещаемое приложением существенно влияет на,

что пользователи считают релевантным. Мы обсудили экспертный поиск выше. Представьте два поисковых приложения медицинской направленности. Оба служат одному кругу пользователей (врачам). Оба хранят одно и то же информационное наполнение (статьи из медицинских журналов). Но имеют одно существенное отличие: одно помогает врачам, лечащим своих пациентов в больницах, а другое позволяет врачам заниматься исследованиями в своих кабинетах. Совершенно разные ожидания предполагают разное понимание релевантности для одного и того же запроса. Поиск по фразе «сердечный приступ» у постели больного должен возвращать действенные и надежные решения страшной проблемы, от которых зависит жизнь человека. В приложении для исследований допускается более широкое разнообразие результатов: по фразе «сердечный приступ» врачи могут искать информацию о новых, интересных исследованиях, не связанную с решением конкретной проблемы.

Часто самое сложное для специалиста по релевантности – понять отношения между контекстом и информационными потребностями. Поисковые запросы пользователей поступают в ваш механизм поиска с большим контекстным багажом. Часть этого багажа приходится на дополнительные данные в виде географических координат или хранящиеся в сеансе пользователя. А другая часть – полностью нематериальная – определяется обещаниями вашего приложения. Приложение создано и продается как инструмент для проведения исследований в спокойной тиши кабинета? Или оно позиционируется как экспертная система, готовая, желающая и способная решить любую задачу, включая помощь врачу, спасающему человеческие жизни?

Рассматривая эти вопросы, можно заметить, что информационный поиск формирует основу для применения обобщенных мер релевантности к чрезвычайно широкому классу проблем. Ваша задача – решить, насколько они применимы к вашему приложению. Как вы увидите далее, многие из них существуют за пределами поисковых технологий и касаются более широких вопросов стратегии: кто ваши пользователи? Что они ожидают от приложения? Какие подразумеваемые и не указанные явно информационные потребности должен удовлетворять поиск?

Фактически, прежде чем двигаться дальше, нужно уточнить наше определение релевантности, чтобы приспособить его для решения прикладной проблемы релевантности:

> *Релевантность* – это практика улучшения результатов поиска *с целью* удовлетворения *информационных потребностей* пользователей в контексте пользовательского опыта и *с учетом потребностей бизнеса*.

1.4. Как решается проблема релевантности?

Получив представление об информационном поиске, сосредоточимся теперь на решении ваших проблем релевантности. Разработчики поисковых механизмов с открытым исходным кодом понимают, что релевантность вашего приложения зависит от широкого спектра факторов. Многие из них характерны для данного конкретного приложения (например, как далеко находится пользователь от того или иного ресторана). Другие являются более универсальными компонентами из информационного поиска.

Учитывая имеющиеся возможности открытых поисковых систем, как бы вы решили прикладную задачу обеспечения релевантности? Как бы вы определили структуру, сочетающую узкоспециальные факторы с более обобщенными приемами информационного поиска?

Для обеспечения релевантности, специалист должен:

1. Выявить существенные *признаки*, описывающие содержимое, пользователей или поисковые запросы.
2. Найти способ сообщить механизму поиска эти признаки через отбор и обогащение.
3. Во время поиска измерять степень релевантности для пользователя, обрабатывая сигналы.
4. Осторожно балансировать влияние нескольких сигналов на результаты, манипулируя функцией ранжирования.

Этот процесс изображен на рис. 1.4.

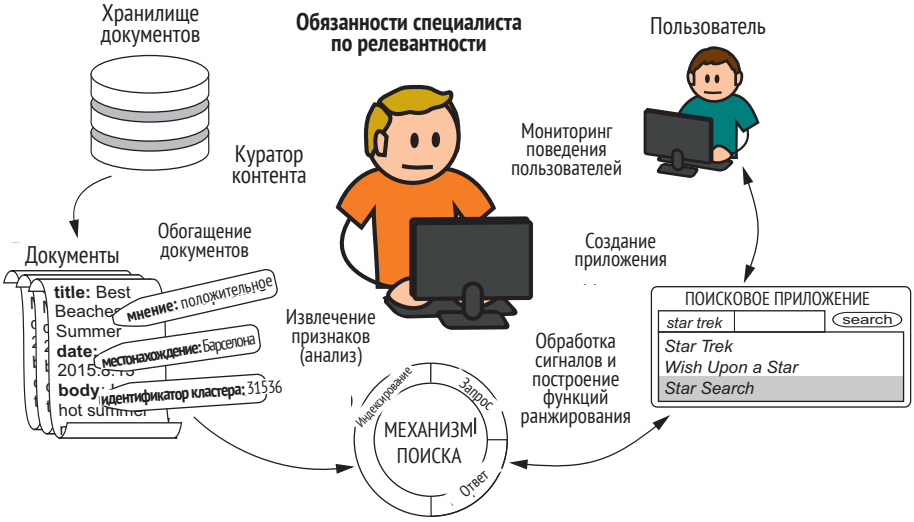


Рис. 1.4. Специалист по релевантности отбирает, обогащает или создает важные признаки и выражает сигналы ранжирования в терминах этих признаков

Звучит немного абстрактно. Что имеется в виду на самом деле? Мы уже обсудили пример выше: как в Google определили признак PageRank

для веб-сайтов (шаг 1). Этот признак оценивается поисковым механизмом Google для каждой веб-страницы (шаг 2). Когда вы выполняете поиск, Google оценивает множество факторов, которые вы считаете важными для релевантности (шаг 3). Например, Google использует алгоритм PageRank непосредственно, как сигнал доверия. К числу других сигналов можно также отнести частоту упоминания искомой вами строки в заголовке/теле страницы, факторы персонализации на основе знания ваших предпочтений. Google смешивает все эти сигналы (шаг 4) в одной большой процедуре ранжирования, которая упорядочивает результаты поиска в надежде, что вы сочтете этот порядок удовлетворительным.

Мы уже обсуждали эти идеи ранее в данной главе. Но давайте попробуем дать более точные определения. *Признак* – это атрибут содержимого или запроса. Признаки управляют принятием решений. Наиболее важной задачей специалиста по релевантности как раз является *выбор признаков* – акт обнаружения и генерации признаков, которые дают соответствующую информацию в процессе поиска.

Знакомые с методиками машинного обучения или классификации могут заметить определенное сходство с признаками. Выполняя классификацию, вы определяете новые признаки данных, чтобы обеспечить более точную классификацию. Данный фрукт – это банан или яблоко? Если известно, что фрукт желтый, велика вероятность, что это банан. Если добавить признак, описывающий форму – круглый или продолговатый – появляется возможность улучшить классификацию. Как видите, признаки помогают поисковым решениям принимать окончательные решения о данных.

Признаки описывают данные, это понятно, но что происходит, когда пользователь выполняет поиск? С помощью сигналов вы командуете механизму поиска ранжировать результаты с использованием ваших определений релевантности. Сигналы измеряют релевантность результатов для данного поиска (конечно же, с использованием признаков!). Например, продолжая пример с механизмом поиска фруктов, пользователь может ввести строку «желтый фрукт». В ответ механизм поиска должен оценить, соответствует ли сорт яблок Голден Делишес (Golden Delicious) запросу этого пользователя. Мы знаем, что цвет имеет значение для покупателей фруктов, поэтому один из сигналов может оценивать, насколько цвет фрукта соответствует цвету, указанному в запросе.

Релевантность редко измеряется одним сигналом. Гораздо чаще ранжирование осуществляется на основе комбинации из нескольких сигналов. Например, кроме цвета покупателя может заботить свежесть продукта. Или, как дополнительный сигнал, пользователь может указать предпочтительный бренд. Мы научим вас управлять функцией ранжирования в механизме поиска так, чтобы результаты выглядели «пугающе умными»,

учитывая все сигналы, которые ваши пользователи могут указать в своих определениях релевантности.

Не волнуйтесь, мы понимаем, что сейчас все эти идеи кажутся вам абстрактными. Когда вы дойдете до следующих глав, к вам начнут приходить те самые озарения «Ага!», помогающие понять, о чем мы сейчас рассказываем. Тем не менее, чтобы получить общее представление, рассмотрим примеры признаков, и как их можно использовать в качестве сигналов на этапе ранжирования:

- *данные о продажах, оценки пользователей* – сообщают о популярных результатах, которые вероятно удовлетворят пользователей;
- *информация о позиции в тексте* – сообщает о совпадении фразы из запроса пользователя с содержимым;
- *текст с синонимами* – сообщает об обнаружении в содержимом синонимов для терминов в запросе;
- *географические координаты* – находится ли что-то близко или далеко: насколько близко находится ищущий к предмету поиска – насколько близко пользователь находится от этого суши-ресторана;
- *признаки для машинного обучения/классификации* – относится ли искомая информация к одному типу (поиск фильмов) или к нескольким (поиск оборудования для ухода за газонами);
- *персонализация/рекомендация* – показал ли пользователь склонность к некоторому определенному типу содержимого? Сможете ли вы отличить других пользователей от выполняющего поиск? Прежние предпочтения пользователя, выполняющего поиск, иногда можно использовать в качестве сигнала, влияющего на результаты поиска.

В следующих главах вы познакомитесь с подходом систематического улучшения релевантности поиска, основанным на отборе признаков и сигналов для ранжирования результатов. Но прежде, чтобы сформировать фундамент для этого знакомства, в главах 2 и 3 мы дадим вам обзор внутреннего устройства механизма поиска и особенности его отладки. В главах с 4 по 7 рассматриваются наиболее насущные проблемы выбора признаков и сигналов. В главе 8 мы представим альтернативные стратегии, помогающие подсказать пользователям путь к нужному им содержимому, когда сам поиск оказывается не в состоянии сделать это.

На протяжении всей книги в качестве основного механизма поиска мы будем использовать Elasticsearch – современный поисковый механизм, основанный на Lucene, широко известной в Java библиотеке функций поиска. Рекомендации, что даются в этой книге, также применимы к Solr, другому поисковому механизму, основанному на Lucene. Хотя наши примеры опираются на Elasticsearch, заложенные в них идеи имеют более широкое применение. Для читателей, использующих Solr, в приложении В описывается соответствие двух поисковых механизмов.

1.5. Не только технологии: кураторство, сотрудничество и обратная связь

Достаточно ли технической основы для решения проблемы релевантности? Вооруженные новыми знаниями, полученными в этой книге, вы, наверное, будете испытывать особое желание улучшить свои поисковые решения. Но ориентация на свои представления о важности – вот главная проблема релевантности. Вы считаете, что даете своим пользователям удивительный опыт поиска и без особой суеты выпускаете обновление – для организации это лишь одна из внутренних, малозаметных задач, которые инженеры просто решают. Это чем-то похоже на выжимание большей производительности из базы данных SQL, правда?

К сожалению, вскоре после выпуска обновленной версии к вам приходит начальник. Ситуация складывается мрачная. Несмотря на все ваши усилия что-то пошло не так. Пользователи не делают покупок. Они не находят нужной информации, прекращают попытки и уходят к конкурентам. Потеряв прибыль, ваш начальник скрипит зубами. В отчаянии он смотрит вам прямо в лицо и требует «сделать поиск более релевантным»! Иными словами, исправить ошибку, реализовать новую функцию, выйти на работу в выходные, если потребуется, но сделать это!

«Сделать более релевантным»? Давайте вспомним определение релевантности. Поразмышляйте над ним и, возможно, тогда вы заметите, какой промах допустили в этой истории:

> *Релевантность* – это практика улучшения результатов поиска с целью удовлетворения *информационных потребностей пользователей* в контексте *пользовательского опыта* и *с учетом потребностей бизнеса.*

Прочитав это определение пару раз, вы быстро заметите, что *специалисты по релевантности не имеют ни малейшего представления, каким должен быть релевантный поиск*! Для удовлетворения информационных потребностей пользователей нужно понимать их цели, предметную область и контекст поиска. Пользователи могут быть самыми разными, от врача, борющегося за жизнь пациента, до бабушки с дедушкой, выбирающих подарок на день рождения внуку или внучке. Чтобы удовлетворить этих пользователей, вы должны мыслить как они. Понимание пользователей далеко выходит за рамки технологий поиска и затрагивает почти все компетенции в организации. Это также верно в отношении понимания потребностей бизнеса, таких как as политика, прибыль, бизнес-цели и другие внутренние факторы.

Для решения проблемы релевантности нужно сместить культуру организации в сторону межфункционального сотрудничества. Как научить специалистов по релевантности понимать язык пользователей и что те

ожидают получить в результате поиска? Как быть, если приложение создается для врачей или юристов? Кто поможет инженеру понять предметную область этих пользователей? Как компания объяснит специалисту по релевантности, что больше всего приносит ей денег? Как осчастливить производителей товаров? Какое содержимое должно занимать первые строчки в результатах поиска (и почему)?

Даже самые обыденные приложения поиска могут испытывать эти сложности. Представьте приложение поиска ресторана. Ваши коллеги по рекламе потрудились привести пользователей к вашему приложению. Теперь приложение, действующее как продавец в магазине (или, может быть, как консьерж?), должно удовлетворить их и заставить вернуться еще раз.

Однако специалист по релевантности не является продавцом. Вводя «суши» в строке поиска, какой ресторан пытается найти пользователь? Принимающий заказы с доставкой на дом? Высокой кухни? Ближайший? Какой-то еще? Другие сотрудники организации, *не* являющиеся специалистами по релевантности, могут понимать, какой цели пытается достичь пользователь. Работать в изоляции для специалиста по релевантности – все равно, что красить дом с завязанными глазами.

Кроме того, сотрудничество с другими сотрудниками не должно ограничиваться простым обучением специалиста по релевантности. *Привлечение к управлению и организации* содержимого для упрощения поиска может оказывать на него не менее благотворное влияние, чем обучение. Вспомните примеры экспертного поиска, приводившиеся выше в этой главе. Здесь опыт библиотекаря по организации информации для упрощения поиска сможет помочь вам усовершенствовать поиск. Часто для этого требуется организовать тесное общение тех, кто глубоко разбирается в информационном наполнении, со специалистами по релевантности, знающими особенности работы поисковых систем.

В основе этих форм сотрудничества лежит идея обратной связи. Эффективная организация стремится обеспечить специалистов по релевантности быстрой и точной обратной связью для информирования и направления их усилий. Графически циклы обратной связи можно изобразить, как последовательность окружностей с уменьшающимися радиусами, как показано на рис. 1.5. Во время первого, самого внешнего цикла, разработчики поиска пребывают в счастливом неведении о значимости релевантности. По мере развития организация переходит к более зрелым формам обратной связи (внутренние окружности): включает данные о поведении пользователей и отзывы экспертов. Наконец, организация преобразует накопленные знания в форму тестов релевантности, обеспечивая практику повышения релевантности через тестирование – наиболее зрелую форму обратной связи.

Рис. 1.5. Формы обратной связи для повышения релевантности поиска

Эта книга в первую очередь учит техническим приемам специалистов по релевантности. Но, размышляя над этими техническими приемами, мы периодически будем напоминать о сказанном выше. Во многих примерах мы явно будем отмечать конкретные результаты, которые хотят видеть пользователи, чтобы вы могли учиться техническим навыкам манипулирования поиском для получения этих результатов. Когда вы будете рассматривать эти примеры, вспоминайте о примерах в этом разделе перед тем, как применять вновь полученные знания для решения своих задач. Более подробно об организационных проблемах мы поговорим в главе 10.

1.6. В заключение

- Релевантность – широко распространенная проблема. Даже в устоявшихся областях, таких как веб-поиск, электронная коммерция и экспертный поиск, продолжается борьба за повышение релевантности результатов.
- Возврат пользователям релевантных результатов поиска может обернуться для бизнеса многомиллионными прибылями; неспособность сделать это может означать проигрыш в конкурентной борьбе.
- *Информационный поиск* – академическая область изучения приемов получения по поисковым запросам содержимого, удовлетворяющего информационные потребности пользователей.
- На практике *релевантность* – это больше, чем удовлетворение информационных потребностей пользователей. Она также призвана удовлетворить потребности бизнеса. Кроме того, понимание инфор-

мационных потребностей пользователей часто зависит от неявной информации, такой как прикладной контекст, цели и опыт.

- Релевантности можно достичь, определив ценные признаки контента и используя их для вычисления сигналов релевантности.
- Технический специалист не сможет добиться хороших результатов в одиночку. Специалист по релевантности часто не имеет навыков оценки степени соответствия (релевантности) содержимого запросам пользователя, опираясь на требования бизнеса, особенности пользовательской аудитории и предметной области.
- Обратная связь имеет жизненно важное значение. С точки зрения специалиста по релевантности, измерение влияния изменений на релевантность помогает избежать возврата неудовлетворительных результатов поиска.